BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 3: e37197
doi: 10.3897/biss.3.37197

OPEN ACCESS

Conference Abstract

Biodiversity Literature Repository (BLR), a repository for FAIR data and publications

Donat Agosti[‡], Terry Catapano[§], Guido Sautter[§], Puneet Kishor[§], Lars Holm Nielsen[|], Alexandros Ioannidis-Pantopikos[|], Chiara Bigarella[|], Teodor Georgiev[¶], Lyubomir Penev[#], Willi Egloff[§]

‡ www.plazi.org, Bern, Switzerland
§ Plazi, Bern, Switzerland
| CERN, Meyrin, Switzerland
¶ Pensoft Publishers, Sofia, Bulgaria
# Pensoft Publishers & Bulgarian Academy of Sciences, Sofia, Bulgaria

Corresponding author: Donat Agosti (agosti@amnh.org)

Received: 12 Jun 2019 | Published: 19 Jun 2019

Citation: Agosti D, Catapano T, Sautter G, Kishor P, Nielsen L, Ioannidis-Pantopikos A, Bigarella C, Georgiev T, Penev L, Egloff W (2019) Biodiversity Literature Repository (BLR), a repository for FAIR data and publications. Biodiversity Information Science and Standards 3: e37197. https://doi.org/10.3897/biss.3.37197

Abstract

Scholarly publications in taxonomy are used as the sole carrier of the communication channel to publicize the description of new species, more generally any kind of taxon, their augmentations in form of re-descriptions to small notes such as additional observation records, or deprecations when the name of a taxon is changing. This is communicated in a highly standardized way. For nomenclatural issues, the Codes (e.g. International Code of Zoological Nomenclature) require certain elements, and for comparative reasons, highly formalized language, document structure, illustrations and citation systems are used. This estimated corpus of 500M printed pages includes all we know about the biodiversity in the sciences as well as what has been processed in terms of specimens in the natural history collections. This includes tens of millions of scientific illustrations, taxonomic treatments, i.e. parts of articles that are clearly delimited and headed by a label including a taxonomic name, materials citation, citations of articles and treatments, specimens by their specimen codes, collections by their collection codes, and collectors by their name. Implicitly, the entire catalogue of life is waiting to be liberated. Only recently, with the advent of text and data mining and of publishing, a change is taking place to make all this data after

This is an open access article distributed under the terms of the CC0 Public Domain Dedication.

publishing immediately widely accessible and digitally accessible knowledge in terms of findability, accessibility, interoperability and reuse (FAIR).

The Biodiversity Literature Repository (BLR) is a community in Zenodo, the open science repository at CERN and part of the European project OpenAIRE. Its goal is to provide a long-term, stable, open access repository for the long tail, preferably FAIR research data. Together with Plazi (http://plazi.org)and Pensoft, BLR has been created to allow deposition of taxonomic articles enhanced with links to data extracted from therein and deposited in BLR, such as illustrations or taxonomic treatments. The metadata for these deposits is enriched with extensive links to cited sources, or the source article. For each of the deposits, a DataCite DOI is minted, unless one exists already. The deposits can be found using either the search functionality of Zenodo, or the BLR website, which makes use of the millions of facts extracted from the publications via the Plazi workflow.

A prime user of the BLR data is GBIF which is using the taxonomic treatments and the figures in their respective view of the taxonomic articles. The input of new taxa allows updating the GBIF taxonomic backbone almost instantaneously and keeping it up-to-date with publishing of new taxa or synonymies.

Currently, 35,600 articles, and 190,000 images are available, with an expected 300,000 taxonomic treatments by Fall 2019 and almost 30% of the annually new described species are accessible online, both in human and machine readable format. With the support of the Arcadia Fund, the goal is to provide access to over 50% of the annually described new species, from the necessary number of taxonomic journals, over 600,000 taxonomic treatments and 350,000 illustrations by 2021, hopefully with an increasing community adding to this new resource as author, publisher, or data liberator.

Keywords

biodiversity informatics, repository, digital literature, open access, taxonomy, taxonomic treatments

Presenting author

Donat Agosti